BESCHWERDEKAMMERN   BOARDS OF APPEAL OF   CHAMBRES DE RECOURS
DES EUROPÄISCHEN   THE EUROPEAN PATENT   DE L'OFFICE EUROPEEN
PATENTAMTS   OFFICE   DES BREVETS

Internal distribution code:
(A) [ ] Publication in OJ
(B) [ ] To Chairmen and Members
(C) [X] To Chairmen
(D) [ ] No distribution

**Datasheet for the decision
of 24 April 2009**

Case Number: T 1416/06 - 3.5.01

Application Number: 00932572.1

Publication Number: 1390869

IPC: G06F 17/30

Language of the proceedings: EN

Title of invention:
Method and system for text mining using multidimensional
subspaces

Applicant:
The Boeing Company

Headword:
Text mining/BOEING

Relevant legal provisions:
EPC Art. 52(2)

Relevant legal provisions (EPC 1973):
EPC Art. 56, 84

Keyword:
"Field of technology"
"Inventive step - main request and auxiliary requests A, B and
D to G (no)"
"Clarity - auxiliary request C (no)"

Decisions cited:
T 0208/84

Catchword:
see point 3

Case Number: T 1416/06 - 3.5.01

D E C I S I O N
of the Technical Board of Appeal 3.5.01
of 24 April 2009

Appellant: The Boeing Company
                                100 North Riverside Plaza
                                Chicago, IL 60606-2016   (US)

Representative: Land, Addick Adrianus Gosling
                                Arnold & Siedsma
                                Sweelinckplein 1
                                2517 GK The Hague   (NL)

Decision under appeal: **Decision of the Examining Division of the
                                European Patent Office posted 8 March 2006
                                refusing European patent application
                                No. 00932572.1 pursuant to Article 97(1) EPC
                                1973.**

Composition of the Board:

Chairman: S. Steinbrener
Members: S. Wibergh
                  P. Schmitz

C1451.D

Summary of Facts and Submissions

I.   This appeal is against the decision of the examining
        division to refuse European patent application
        No. 00932572.1.

II.   The following document will be referred to:

        D2: M.W. Berry et al., "Low-Rank Orthogonal
             Decompositions for Information Retrieval
             Applications", Numerical Linear Algebra with
             Applications, I(1), 1996, 1-27.

III.   According to the decision appealed the then main and
        first auxiliary requests contained undisclosed subject-
        matter. The invention according to the second auxiliary
        request was not inventive, and claim 1 of the third
        auxiliary request was not clear.

IV.   With the statement setting out the grounds of appeal
        dated 17 July 2006, the appellant requested that the
        decision be set aside and a patent be granted based on
        the claims according to the main request or auxiliary
        requests A to G, all filed with the same letter. The
        main request and auxiliary request C corresponded in
        essence to the second and third auxiliary requests
        before the examining division, respectively. Auxiliary
        requests D-G were filed as a matter of precaution in
        case the preceding requests were not found to meet the
        requirements of Article 123(2) EPC.

V.   In a communication accompanying the summons to oral
        proceedings requested by the appellant as an auxiliary
        measure, the Board stated that it had doubts whether

the characterising features solved a technical problem. Providing "verbal semantics", ie "words or terms" for describing the meaning of a dimension seemed basically to be a presentation of information. The verbal semantics was directed to the mind of a user. It was therefore not clear that there was a technical effect indicating the presence of technical means for solving a technical problem. As to the auxiliary requests the Board stated that it was not apparent that any part of the invention according to the independent claims of these requests contributed to the solution of a technical problem. The steps were of a mathematical nature, serving to process information with the purpose of improving its presentation or visualisation to the user. Whether or not this was new and original could not play any role as long as no technical effect was achieved. As to the skilled person it was clear that mathematical skills were crucial. A mathematician was however not a technically skilled person but could at most be part of a technical team.

VI.   Oral proceedings were held on 24 April 2009. In the early hours of the same day the representative had sent a fax stating that the appellant had decided not to attend the oral proceedings. The oral proceedings were held in the appellant's absence.

VII.   The Board verified that the appellant had requested in writing that the decision under appeal be set aside and a patent be granted on the basis of the main request or auxiliary requests A to C, filed with the statement setting out the grounds of appeal, or on the basis of auxiliary requests D to G corresponding to the foregoing requests but including in the independent

claims all the features of dependent claims 5-12 as
originally filed.

VIII. Claim 1 according to the *main request* reads:

"Method of representing a document collection in a
computer system, wherein the document collection
comprises a plurality of documents, with each document
comprising a plurality of terms, using a matrix
decomposition based on a distribution of the frequency
of occurrences of each of the terms in each of the
documents, the method comprising:
(a) constructing a term frequency matrix, having a
dimension represented by the documents and another
dimension represented by the terms, wherein each
element of the term frequency matrix is the frequency
of occurrence of one of the terms in one of the
documents;
(b) determining a projection type;
(c) determining a lower dimensional subspace;
(d) determining a number of matrix dimensions to use;
(e) generating an original term subspace by projecting
the projection type into the lower dimensional subspace,
wherein the step of projecting comprises performing a
truncated two-sided orthogonal decomposition of the
term frequency matrix according to the determined
number of dimensions, the decomposition identifying
significant features using three matrices: a term basis
matrix (U_c), a weight matrix (R_c, L_c) and a document
basis matrix (V_c), wherein the term basis matrix and the
document basis matrix both have orthonormal columns;
characterized by

(f) providing verbal semantics for at least one
dimension of the subspace, wherein providing verbal
semantics for the dimension comprises:
(f1) identifying a column in the term basis matrix that
represents a desired dimension;
(f2) identifying a plurality of terms in the column
with the largest absolute values for positive elements;
(f3) identifying a plurality of terms in the column
with the largest absolute values for negative elements;
and
(f4) returning the identified positive elements and the
identified negative elements as a contrast set
describing the meaning of position along the dimension".

IX.   According to *auxiliary request A* the preamble of
        claim 1 is the same as for the main request whereas the
        characterising part reads:

        "characterized by
        (f) providing a legend for at least one dimension of
        the subspace, wherein providing a legend comprises:
        (f1) identifying a column in the term basis matrix {or
        term frequency matrix ?} [*sic*] that represents a
        desired dimension;
        (f2) identifying a plurality of terms in the column
        with the largest absolute values for positive elements;
        (f3) identifying a plurality of terms in the column
        with the largest absolute values for negative elements;
        and
        (f4) returning the terms with the identified positive
        elements and the identified negative elements;
        characterized by [*sic*]
        (f5) generating a legend for the dimension, the legend
        consisting of one or more terms of the identified

positive elements attached to one direction of the
dimension and one or more terms of the identified
negative elements attached to the other direction of
the dimension".

X.   According to *auxiliary request B* the following features
          are added to claim 1 of the main request:

          "(g) performing information visualization, wherein
          performing information visualization comprises:
          (g1) in response to a user request for dimensions of
          visualization, determining a number of dimensions;
          (g2) computing the requested dimensions;
          (g3) determining whether the user has requested fewer
          than the determined number of dimensions;
          (g4) if the user has requested fewer than the
          determined number of dimensions, using a set of default
          dimensions;
          (g5) determining if the dimensions are orthogonalized;
          (h) if the dimensions are not orthogonal, taking an
          appropriate action based on a user preference, wherein
          the user preference comprises one of:
                 (i) orthogonalizing the dimensions; and
                 (ii) providing an indication that the dimensions
                 are not orthogonal;
          (i) projecting the documents onto the dimensions;
          (j) generating labels for the dimensions from the
          returned positive and negative elements;
          (k) displaying a plurality of indicators, wherein each
          indicator represents a document, on a plurality of
          labeled axes corresponding to the dimensions".

XI.   According to *auxiliary request C* the preamble of
         claim 1 is the same as for the main request whereas the
         characterising part reads:

         "characterized by
         (f) providing verbal semantics for at least one
         dimension of the subspace;
         (g) updating the original term subspace as the document
         collection changes, wherein said updating the original
         term subspace comprises:
         - identifying a plurality of new documents;
         - identifying a plurality of new terms in the new
         documents;
         - constructing a new term frequency matrix that
         represents the new documents;
         — projecting the new term frequency matrix on the
         original term subspace, the term basis matrix;
         — computing a residual;
         - augmenting the existing term subspace with the
         residual;
         - expanding the original term subspace, wherein
         expanding the original term subspace comprises
         expanding the document basis matrix by adding a small
         identity matrix; and
         — re—identifying significant features in the
         subspace".

XII.   No sets of claims according to auxiliary requests D-G
         were filed. The appellant merely indicated in the
         statement setting out the grounds of appeal that the
         claims according to these requests were to correspond
         to the claims of the main request and auxiliary
         requests A-C, respectively, with the independent claims

additionally including all features of claims 5-12 as originally filed.

XIII. At the end of the oral proceedings the Board announced its decision.

Reasons for the Decision

All requests

1.   Construction of claim 1

Claim 1 is directed to a method of "representing a document collection in a computer system". On the basis of the description (see eg p. 10, l. 10-14 and fig. 2) this formulation is taken to mean that the method is performed by means of a computer system, not that the collection is stored in a computer system.

2.   Exclusion under Article 52(2) EPC

With the above interpretation of claim 1 the method is an invention within the meaning of Article 52(1) EPC since it comprises a computer system.

3.   Field of technology

3.1   Claim 1 of all requests is directed to a method of representing a document collection. The method is to a large extent defined in terms of equations. The purpose of the method is to present the information in a way that can be more easily understood or evaluated by a user. A fundamental question in this context is whether

C1451.D

the invention - apart from its being implemented on a
computer - is within a field of technology. At the
bottom of the method is a mathematical technique known
as orthogonal decomposition. This technique is
generally applied to large matrices and, like many
mathematical functions, can be represented graphically.
It is typical for mathematical representations that
they involve pure numbers, ie abstract data, having no
physical connotation. In the present invention the
representations are of documents and the terms used in
the documents. Thus, although the data have a certain
"meaning", they remain abstract. They can hardly be
regarded as forming a physical entity, nor does the
method result in a change in the data but merely in
their representation (cf T 208/84 "Computer-related
invention/VICOM", OJ EPO 1987,14, point 5 of the
Reasons). It could therefore be argued that the
invention - again apart from its implementation - is
essentially a mathematical method pursuant to
Article 52(2)(a) EPC, resulting in a presentation of
information pursuant to Article 52(2)(d) EPC.

3.2   In the following these general concerns will however
        not be pursued since there are more specific reasons
        for not allowing the appeal.

Main request

4.   The prior art

        The appellant acknowledges (statement setting out the
        grounds of appeal, point 4.2, first paragraph) that D2
        discloses the preamble of claim 1. The examining
        division held that in addition it implicitly disclosed

feature (f) of providing verbal semantics except for the identification of a plurality of terms in the column with the largest absolute values for positive and negative values (decision under appeal, point 3.3). The Board agrees with this view since in the diagram shown in fig. 4 of D2 all terms are indicated, implying that the corresponding columns have been identified and the respective elements returned.

5.   Novelty

According to the invention a plurality of terms in the column with the largest absolute values for positive and negative values are identified. These terms are used to describe the meaning of respective dimensions. In D2 *all* terms are identified, not just the largest ones. It could be argued that the claim feature is not actually a distinction since according to D2 the largest values are indeed identified (as a sub-group of all values). However, to the benefit of the appellant it is assumed that a means for identifying the largest values (and discarding smaller ones) is implied. This would be a new feature (Article 54(1)(2) EPC 1973).

6.   Inventive step

The invention serves to indicate to the user what terms a document is likely to contain (see the paragraph bridging pp. 21 and 22 of the description). The solution concerns the legends (labels consisting of one or more words) describing the axes (dimensions) so that documents near one end of the dimension tend to contain the words at that end or words correlated with those words in the document set. As most people are aware,

diagram axes are usually labelled. The question in the
present case is therefore whether it was obvious to
label the axes with the most dominant term or terms.
Apart from any personal preferences of the users the
answer appears to depend on the size of the database. A
data base visualisation may comprise several hundred
dimensions (as indicated on the first page of D2). The
axis labels could not indicate all associated terms but
a selection must be performed, and common sense
dictates that the dominant terms should be selected.
Thus the subject-matter of claim 1 does not involve an
inventive step (Article 56 EPC 1973)).

Auxiliary request A

7.   Inventive step

Ignoring the comment "{or term frequency matrix ?}" in
claim 1, the Board finds that the only difference
between auxiliary request A and the main request is
that a legend containing the axis information is
generated. This feature has already been considered
above. Thus also this subject-matter does not involve
an inventive step (Article 56 EPC 1973).

Auxiliary request B

8.   Clarity, support in the description

Claim 1 has been supplemented by features of original
claim 23 including the features "in response to a user
request for dimensions of visualization, determining a
number of dimensions" and "determining whether the user

has requested fewer than the determined number of dimensions". These features are contradictory.

The claim also reiterates from claim 23 the feature that if a user has requested fewer than the determined number of dimensions, a set of default dimensions is used. This could be taken to mean that the default dimensions (entirely) replace the requested dimensions. On p. 21, l. 11-13 of the description, however, it is explained that the user request is "filled out with default axes", not that it is replaced.

Therefore the claim is not clear and supported by the description in the sense of Article 84 EPC 1973. In the following it will be interpreted in accordance with the description.

9.   Inventive step

The additional features in claim 1 permit the user himself to define dimensions of visualisation. If necessary the dimensions are orthogonalised. Axis labels are generated.

The wish to be able to study particular aspects (terms) of the data collection follows directly from a user's particular interests, which are known to be subjective and therefore cannot render the problem formulation inventive. Given that such interests might exist, it was clear that the system should be designed such that the displayed axes are selectable. This implies choosing a particular subspace of the total term space, ie a particular projection, a technique that is well understood. The choice offered to the user between

orthogonalisation or no orthogonalisation merely takes
into account the fact that different users have
different preferences.

Thus the subject-matter of claim does not involve an
inventive step (Article 56 EPC 1973), and the request
is refused.

Auxiliary request C

10.   Clarity

The examining division decided that the features "small
identity matrix" and "significant features" were not
clear (decision under appeal, point 4.1). The Board
agrees. In addition, the feature "computing a residual"
leaves open what this residual is and how it is
computed. The corresponding part of the description
(fig. 7 and p. 19, top) contains no equations and no
further explanations. Also this feature must thus be
regarded as not clear (Article 84 EPC 1973). Therefore
auxiliary request C is refused.

Auxiliary requests D-G

11.   These requests correspond to the four foregoing
requests but include in the independent claims all the
features of dependent claims 5-12 as originally filed.
The appellant filed these requests in case the Board
would be of the opinion that the foregoing requests
were not allowable under Article 123(2) EPC. It was not
argued that the additional features rendered the
claimed subject-matter inventive, nor is this apparent

to the Board. Thus these four last requests are also refused (Article 56 EPC 1973).

Non-attendence at oral proceedings

12.   The appellant's fax indicating his intention not to attend the oral proceedings was received by the EPO at 4.29 hrs on the day of the oral proceedings so that the Board could not have been made aware of it before the oral proceedings were opened. The Board therefore had to wait and see whether somebody representing the appellant might have been delayed and eventually had to ask the registrar of the Board to make inquiries by phone calls. This situation is clearly undesirable and should be avoided by providing information of non-attendance in due time.

Order

For these reasons it is decided that:

The appeal is dismissed.

The Registrar: The Chairman:

T. Buschek   S. Steinbrener